Kalaiselvan

Create ETL Project for Collecting Sales Data using SSIS

Case Scenario

For this ETL project, the requirements are listed below:

- Sales data will be pushed to specified shared folder regularly.
- Data is stored in CSV files with columns: Sale Number, Product Name, Product Color, Sale Amount, Sale Area and date.
- Minions of sales records are stored in each file.
- The flat files need to be archived after processing.
- The process can be scheduled to run automatically and periodically.

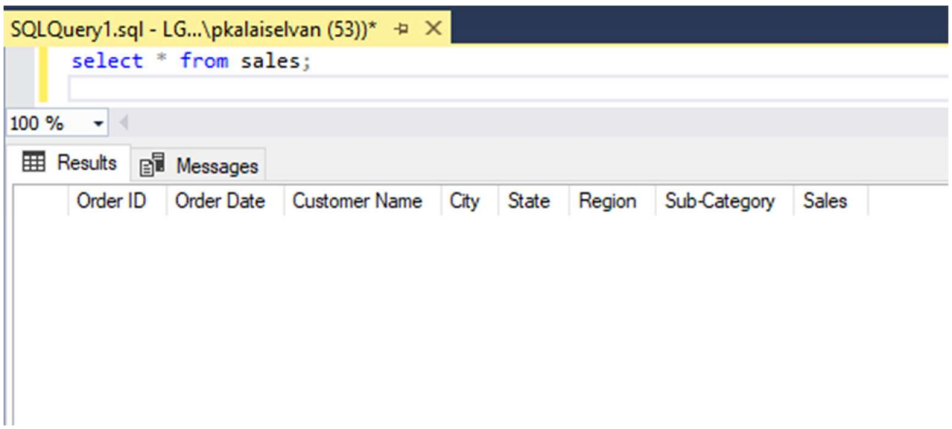
For datasets:

https://data.world/datasets/sales

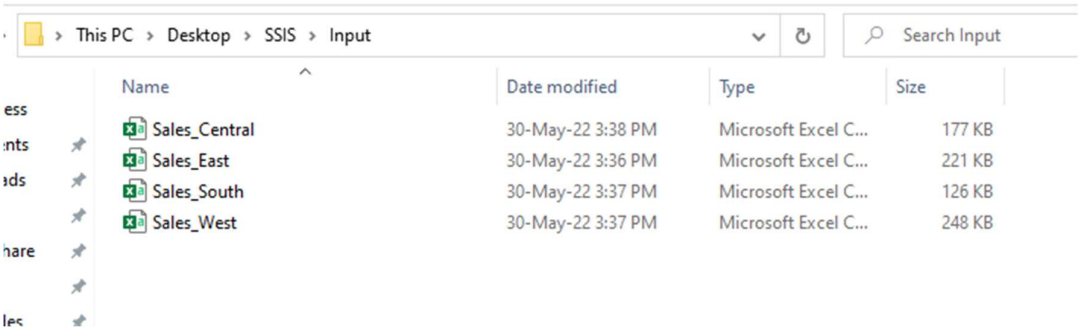
Steps Followed:

1. Created table in SQL Server Management Studio
2. Placed .csv data files in to 'Input' folder
3. Created SSIS package that fetch data from .csv files (from input folder) and load data to SQL server table (sales)
4. After data load, those .csv files will be moved to 'Archive' folder with timestamp
5. Created SQL Server Agent - job and schedule to automate this process.

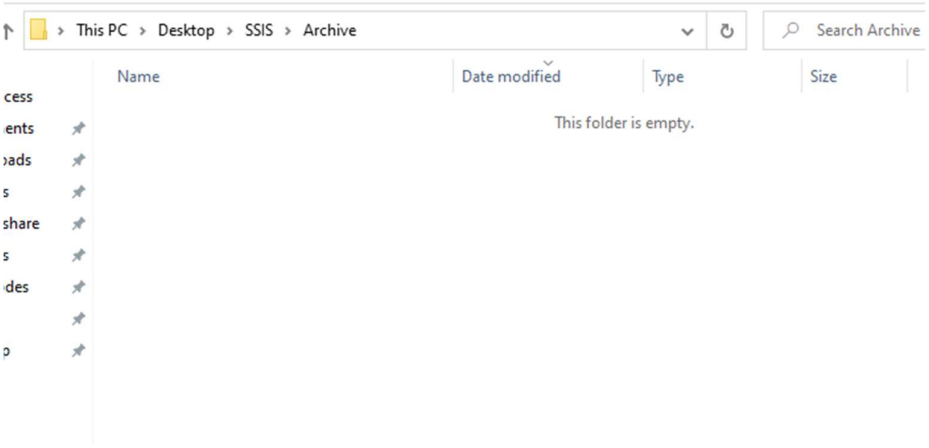
1. Created table in SQL server (sales)



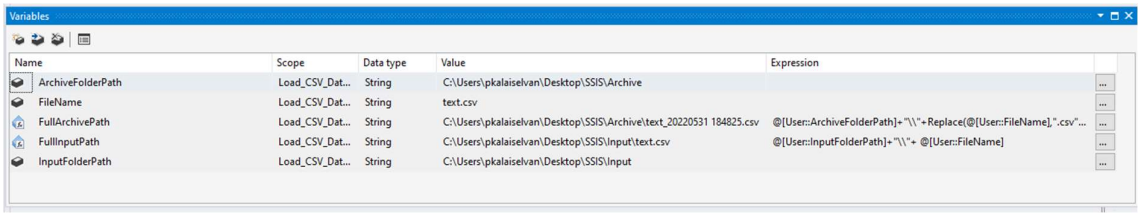
2. Placed .csv files in 'Input' folder



3. Archive folder is empty now, files will be saved in this folder after processing the data load

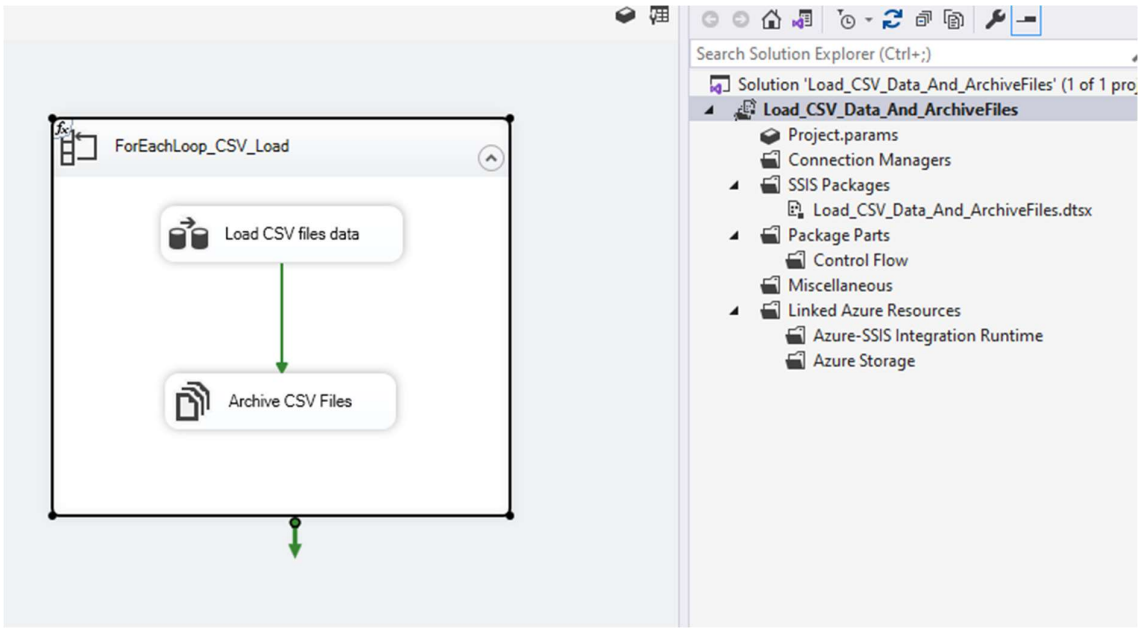


4. Created variable that holder folder path, filename and full path of the folder

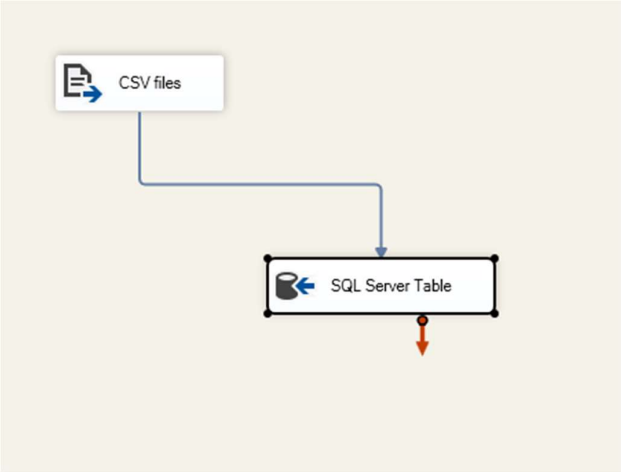


5. Created SSIS package that loads data from .csv files to SQL server
    a. Used 'ForEachLoop Container' to process multiple .csv files
    b. Created 'Data Flow Task' the fetch all the .csv files from 'Input' folder and loads data to SQL server table
    c. Used 'File System Task' to move the .csv files in to 'Archive' folder with processed timestamp.

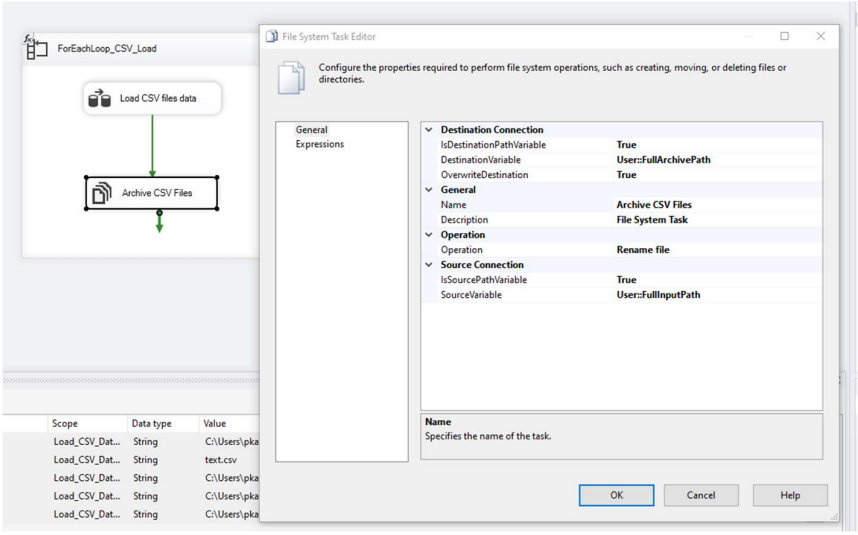
SSIS Package:



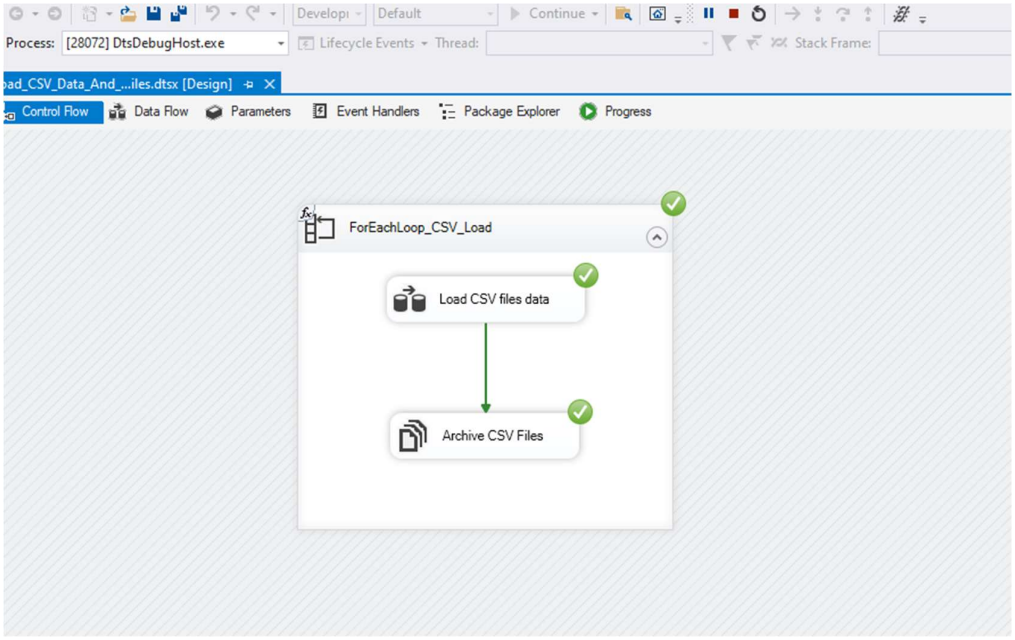
Data Flow Task



File System Task



6. After the successful load, data are available in SQL server table



SQLQuery1.sql - LG...\pkalaiselvan (53))*

```
select * from sales;
```

100 %

Results | Messages

	Order ID	Order Date	Customer Name	City	State	Region	Sub-Category	Sales
1	US-2012-118983	22-11-13	Harold Pawlan	Fort Worth	Texas	Central	Appliances	68.81
2	US-2012-118983	22-11-13	Harold Pawlan	Fort Worth	Texas	Central	Binders	2.54
3	CA-2011-105893	11-11-12	Pete Kriz	Madison	Wisconsin	Central	Storage	665.88
4	CA-2013-137330	10-12-14	Ken Black	Fremont	Nebraska	Central	Art	19.46
5	CA-2013-137330	10-12-14	Ken Black	Fremont	Nebraska	Central	Appliances	60.34
6	CA-2014-107727	20-10-15	Matt Abelman	Houston	Texas	Central	Paper	29.47
7	CA-2013-117590	09-12-14	Gene Hale	Richardson	Texas	Central	Phones	1097.54
8	CA-2013-117590	09-12-14	Gene Hale	Richardson	Texas	Central	Furnishings	190.92
9	CA-2012-117415	27-12-13	Steve Nguyen	Houston	Texas	Central	Envelopes	113.33
10	CA-2012-117415	27-12-13	Steve Nguyen	Houston	Texas	Central	Bookcases	532.4
11	CA-2012-117415	27-12-13	Steve Nguyen	Houston	Texas	Central	Chairs	212.06
12	CA-2012-117415	27-12-13	Steve Nguyen	Houston	Texas	Central	Phones	371.17
13	CA-2014-120999	11-09-15	Linda Cazamias	Naperville	Illinois	Central	Phones	147.17
14	CA-2013-118255	12-03-14	Odella Nelson	Eagan	Minnesota	Central	Accessories	45.98
15	CA-2013-118255	12-03-14	Odella Nelson	Eagan	Minnesota	Central	Binders	17.46
16	CA-2011-146703	20-10-12	Patrick O'Donnell	Westland	Michigan	Central	Storage	211.96
17	CA-2012-115742	18-04-13	Darren Powers	New Albany	Indiana	Central	Binders	38.22
18	CA-2012-115742	18-04-13	Darren Powers	New Albany	Indiana	Central	Labels	75.18
19	CA-2012-115742	18-04-13	Darren Powers	New Albany	Indiana	Central	Furnishings	6.16
20	CA-2012-115742	18-04-13	Darren Powers	New Albany	Indiana	Central	Chairs	89.99
21	US-2012-164175	30-04-13	Paul Stevenson	Chicago	Illinois	Central	Chairs	213.12
22	CA-2014-114440	15-09-15	Tracy Blumstein	Jackson	Michigan	Central	Paper	19.05
23	US-2014-118038	10-12-15	Ken Brennan	Houston	Texas	Central	Binders	1.25

7. Once the data load is completed, the files are moved from ‘Input’ folder to ‘Archive’ folder with date and timestamps